Adapting an Automatic Speech Recognition System to Event Classification of Electroencephalograms[1]

V. Shah, R. Anstotz, I. Obeid and J. Picone

The Neural Engineering Data Consortium, Temple University
{vinitshah, ryan.anstotz, iobeid, picone}@temple.edu

Identification of clinically significant events in electroencephalograms (EEGs) is a time-consuming task for neurologists [1]. EEG signals contain a variety of morphologies which relate to a combination of brain signals and noise/artifacts. Automated classification of such events has the potential to speed up the interpretation process and provide valuable input to other types of EEG decision-making software. Because of the similarities between EEGs and speech signals, both of which contain temporal/sequential information, one of our long-term goals has been to apply well-developed concepts from speech recognition to EEG processing. We have previously approached this by applying hidden Markov Models (HMMs) [2][3] using a toolkit known as HTK [4]. In this poster, we discuss the application of a new high-performance speech recognition system known as Kaldi [5] to this task. Adaptation of this technology to the EEG problem has not been as straightforward as previously thought.

Kaldi is an extremely popular open source toolkit that integrates many types of relatively new deep learning algorithms with more traditional HMM approaches. Though it is designed to be flexible, configuring it to complete non-speech recognition related tasks requires substantial modifications to the way the software handles sequential data. In this study, we adapt Kaldi to do EEG event classification on six types of EEG events: periodic lateralized epileptiform discharges (PLED), generalized periodic epileptiform discharges (GPED), spike/sharp and wave discharges (SPSW), eye movements (EYEM), artifacts (ARTF), and background (BCKG). The first three events are of clinical interest [6]. The last three events are used to model various types of background noise. We have developed a database, known as the TUH EEG Events Corpus (TUEC), that can be used to model these events [7] and have reported classification results for a number of algorithms [3]. In this study, we have developed systems based on Kaldi and compared performance to our previous approaches.

Classification is performed using a 26-dimensional feature vector consisting of Linear Frequency Cepstral Coefficient (LFCC) features which were captured from the EEG signals. The feature vector contains energy, the first seven cepstral coefficients, and the first and second derivatives of the cepstral coefficients [8]. The HMM topology for each event is the same – a 3-state Bakis model [2]. We use Gaussian Mixture Models (GMMs) for output distributions at each state in each HMM. During acoustic modeling, Kaldi HMMs are modeled based on pdf-ids [5]. Pdf-ids are GMM indices associated with individual probability density functions (PDFs). They are extracted from the context dependent decision trees where leaves of the tree represent the pdf-ids. We use 40 iterations of Viterbi training to estimate the parameters of the HMMs. A diagonal covariance matrix assumption is used at each state. Tuning experiments determined that a total 50 Gaussian mixture components were optimal.

We also evaluated the application of an adaptation technique known as Maximum Likelihood Linear Transforms (MLLTs) which performs adaptation on top of the transformed features (on pdf-ids) via Linear Discriminant Analysis (LDA). MLLT, also known as Semi-Tied Covariance (STC) [9][10], is a model-state transformation technique to estimate a global covariance matrix which allows a limited number of full covariance matrices to be shared over GMM distributions. This helps the system model correlations among features at a very low computational cost. We use 35 iterations of Viterbi training while intermittently updating the MLLT transformation matrix four times.

1. Research reported in this publication was most recently supported by the National Human Genome Research Institute of the National Institutes of Health under award number U01HG008468. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Finally, we developed a system based on deep neural networks [11] and HMMs, referred to as DNN-HMM, by replacing GMMs with Multi-Layer Perceptrons (MLPs) to model the observation distribution. The deep network consists of three hidden layers with 256 neurons per layer with rectified linear units (ReLU) as activation functions. The output layer contains six neurons (for each class) with Softmax activation function. The system is trained using a Stochastic Gradient Descent (SGD) optimizer and an annealing learning rate after each epoch.

The Viterbi decoding algorithm [12] is used to calculate the probability of observing the sequences and output of each utterance stored in the form of a lattice [13]. Lattices contain outputs of N-best set of hypotheses of phone/word sequences. Each node in the Kaldi-lattice includes acoustic and language model scores along with time information. Since the number of events to be evaluated is only six, we kept the lattice-beam value low (0.7 – 1.0) during decoding.

Table 1 shows the performance of our baseline GMM-HMM system implemented using HTK with a total of 12,494 HMM parameters. Similarly, Table 2 provides the performance of a comparable GMM-HMM baseline implemented using Kaldi that uses 8,438 parameters. The Kaldi HMM's Viterbi training requires ~40 minutes to train the models on 8 CPU cores whereas HTK HMMs use the Baum-Welch reestimation algorithm and require same amount of time using only 1 CPU core for training. Both these systems were scored and compared using the Epoch scoring metric [14]. Similarly, Table 3 and Table 4 compare performance of the LDA-MLLT and MLP-HMM systems, respectively. Kaldi's LDA-MLLT system performs better than its other variants with an average detection rate of 37.42%, but still underperforms compared to HTK baseline system (57.31%). All of the Kaldi HMM variants perform very poorly on SPSW detection, since they are mainly misclassified with the GPED or BCKG events.

This study suggests that EEGs possess similar behavior to that of speech waveforms. So, speech recognition tools such as Kaldi ASR and HTK, which perform temporal/sequential classification, can be directly adapted for EEG event classification. The Kaldi HMMs developed for the six-event classification does not show any improvement in performance compared to HTK baseline system. LDA-MLLT system's overall performance is better than its other variants but the systems, which use LDA features, perform extremely poorly on SPSW events.

Ref/Hyp	BCKG	EYEM	ARTF	PLED	GPED	SPSW
BCKG	71.93	2.59	7.02	2.28	7.37	8.81
EYEM	0.61	82.37	2.13	8.51	2.13	4.26
ARTF	45.19	2.18	41.24	2.77	3.81	4.81
PLED	1.85	4.70	0.70	54.80	17.62	20.32
GPED	4.85	2.39	7.46	20.42	53.32	11.55
SPSW	8.29	9.17	4.41	4.59	33.33	40.21

Table 1. Performance of the baseline GMM-HMM (HTK) system

Ref/Hyp	BCKG	EYEM	ARTF	PLED	GPED	SPSW
BCKG	67.68	10.78	1.93	5.38	5.26	8.97
EYEM	42.09	48.90	5.62	2.80	0.26	0.33
ARTF	41.30	40.19	13.76	2.59	1.41	0.76
PLED	3.94	7.39	1.58	46.68	18.19	22.22
GPED	11.14	12.35	6.21	41.00	20.36	8.94
SPSW	28.62	16.29	2.95	6.46	34.26	11.41

Table 2. Performance of the baseline GMM-HMM (Kaldi) system

Ref/Hyp	BCKG	EYEM	ARTF	PLED	GPED	SPSW
BCKG	62.08	8.69	15.21	2.29	1.52	10.21
EYEM	34.89	54.36	8.60	0.22	1.92	0.00
ARTF	27.03	38.43	30.67	0.63	2.52	0.72
PLED	2.26	11.38	5.31	43.08	26.87	11.11
GPED	18.71	3.98	10.58	29.80	31.01	5.92
SPSW	17.88	18.64	21.69	1.34	37.12	3.33

Table 3. Performance of an LDA-MLLT system

Ref/Hyp	BCKG	EYEM	ARTF	PLED	GPED	SPSW
BCKG	73.79	2.71	13.31	0.35	2.18	7.63
EYEM	68.11	25.75	1.70	0.42	0.35	3.63
ARTF	38.65	27.87	24.35	0.39	2.58	6.15
PLED	7.70	3.52	3.96	46.83	28.38	9.58
GPED	10.80	0.39	43.78	21.20	20.87	2.93
SPSW	42.04	9.7	9.81	0.04	36.55	1.80

Table 4. Performance of a Kaldi-based DNN-HMM system

REFERENCES

[1] K. A. Jellinger, "Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields, 6th edn," *Eur. J. Neurol.*, 2011.

[2] J. Picone, "Continuous Speech Recognition Using Hidden Markov Models," *IEEE ASSP Mag.*, vol. 7, no. 3, pp. 26–41, Jul. 1990.

[3] M. Golmohammadi, A. H. H. N. Torbati, S. Lopez, I. Obeid, and J. Picone, "Automatic Analysis of EEGs Using Big Data and Hybrid Deep Learning Architectures," *J. Clin. Neurophysiol.*, pp. 1–30, 2018.

[4] "HTK," *Machine Intelligence Laboratory, Department of Engineering, Cambridge University*, 2009. [Online]. Available: http://htk.eng.cam.ac.uk/.

[5] D. Povey *et al.*, "The Kaldi speech recognition toolkit," in *IEEE Workshop on Automatic Speech Recognition and Understanding*, 2011, pp. 1–4.

[6] G. L. Krauss and R. S. Fisher, *The Johns Hopkins Atlas of Digital EEG: An Interactive Training Guide*. Johns Hopkins University Press, 2006.

[7] I. Obeid and J. Picone, "The Temple University Hospital EEG Data Corpus," *Front. Neurosci. Sect. Neural Technol.*, vol. 10, p. 196, 2016.

[8] A. Harati, M. Golmohammadi, S. Lopez, I. Obeid, and J. Picone, "Improved EEG Event Classification Using Differential Energy," in *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium*, 2015, pp. 1–4.

[9] M. J. F. Gales, "Maximum likelihood linear transformations for HMM-based speech recognition," *Comput. Speech Lang.*, vol. 12, no. 2, pp. 75–98, 1998.

[10] M. J. F. Gales, "Semi-tied covariance matrices for Hidden Markov Models," *Speech Audio Process. IEEE Trans.*, vol. 7, no. 3, pp. 272–281, 1999.

[11] D. Povey, X. Zhang, and S. Khudanpur, "Parallel training of DNNs with Natural Gradient and Parameter Averaging," in *International Conference on Learning Representations (ICLR)*, 2015, p. 16.

[12] A. Viterbi, "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm," *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 260–269, Apr. 1967.

[13] F. Richardson, M. Ostendorf, and J. R. Rohlicek, “Lattice-based search strategies for large vocabulary speech recognition,” in *1995 International Conference on Acoustics, Speech, and Signal Processing*, 1995, pp. 576–579.

[14] V. Shah, S. Ziyabari, M. Golmohammadi, I. Obeid, and J. Picone, “Objective Evaluation Metrics for Automatic Classification of EEG Events,” *J. Neural Eng.*, pp. 1–19, 2018 (in review). *https://www.isip.piconepress.com/publications/unpublished/journals/2018/iop_jne/metrics/*.

NEURAL ENGINEERING DATA CONSORTIUM
www.nedcdata.org

Adapting an Automatic Speech Recognition System to Event Classification of EEGs

V. Shah, R. Anstotz, I. Obeid and J. Picone
The Neural Engineering Data Consortium, Temple University

College of Engineering
Temple University

Abstract

- Interpretation of electroencephalogram (EEG) events is a tedious, time-consuming and expensive task.
- Automatic interpretation will accelerate the review process and lead to better healthcare outcomes.
- In this study, we analyze EEGs in terms of six types events which are either of clinical interest or are related background noise.
- Due to the similarities between speech and EEG signals, machine learning (ML) algorithms developed for automatic speech recognition (ASR) can be used for identification of the six-way events.
- We adapt a very well-known state of the art automatic speech recognition (ASR) toolkit, Kaldi.
- We developed a Multi-pass Kaldi system that integrates a hidden Markov model (HMM) based system for segmentation, a maximum likelihood linear transformation (MLLT) system for adaptation and a multilayer perceptron (MLP) deep learning system for classification.
- Unfortunately, Kaldi delivers lower performance (37.5% sensitivity) than our previous best HMM system implemented using HTK (57.3%).

TUH EEG Events Corpus (TUEV–v1.0.1)

Contains annotations for three types of clinically-relevant EEG events:

- Periodic Lateralized Epileptiform Discharge (PLED)
- Generalized Periodic Epileptiform Discharges (GPED)
- Spike or Sharp and Wave discharges (SPSW)

and three types of events used to model background:

- Artifacts (ARTF)
- Eye Movements (EYEM)
- Background (BCKG)

	Train	Eval
Patients / Sessions[1]	290 / 290	80 / 80
Duration (Hrs)	100.5	48.3

Note: By design there is one session per patient.

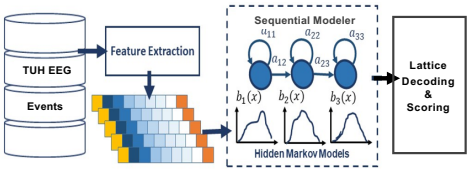
- The release includes signal data stored in an EDF format, per-channel annotations of the six events (.lab format) , HTK formatted features, and an associated EEG report for the session.
- The proportion of the partially annotated EEG events is fairly balanced between the train and eval sets.

Events	Train Set		Eval Set	
	Epochs	Dist. (%)	Epochs	Dist. (%)
PLED	11,254	13.4	4,677	15.9
GPED	6,184	7.4	1,998	6.8
SPSW	645	0.8	567	2.0
ARTF	11,053	13.2	2,204	7.5
EYEM	1,170	1.4	329	1.1
BCKG	53,726	63.9	19,646	66.8
Total	84,032	100.0	29,421	100.0

Adapting Kaldi to EEG Interpretation

Background:

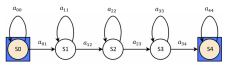
- HTK is a portable toolkit for building and manipulating Hidden Markov Models (HMMs).
- Kaldi is a similar toolkit that has gained widespread adoption due to its state of the art performance on large vocabulary speech recognition tasks.
- Kaldi is based on finite state transducer technology and integrates powerful deep learning technology.



System Overview:

- Feature Extraction: Features (26) consist of energy terms (2), linear frequency cepstral coefficients (7), deltas (9) and delta-deltas (8).
- Acoustic Model: A simple 3-state left-to-right Bakis model topology with Gaussian Mixture Models (GMMs) for output distributions (8).
- Language Model: Since EEG events do not appear in a specific order, no LM penalties were applied. However, postprocessing is typically used in a later stage.
- Training: The Viterbi algorithm is used to estimate the parameters of the HMMs (40 iterations).

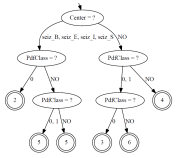
Bakis Model:



- Traditionally used in speech recognition.
- Enforces a dynamic time-warping of the signal.
- Each event uses the same number of states.

Decision Trees:

- Ties HMM states based on acoustic phonetic features such as tone and stress.
- Allows GMM components to be efficiently shared.



Performance on Six-Way Classification (TUEV)

Models:

- GMM-HMM Monophone System: This is a flat-start model and was trained using Viterbi algorithm for 40 iterations with a total of 50 gaussian components.
- LDA-MLLT Triphone System: Used MLLT to estimate a global covariance matrix to be shared over GMM distributions. Used Linear Discriminant Analysis (LDA) to reduce the dimensionality of the classes.
- DNN-HMM MLP System: Used a deep neural network (DNN) to model observation distributions instead of GMMs. Fully connected layers with 256 neurons are used with 3 hidden layers. A Rectified Linear Unit (ReLU) activation function and Stochastic Gradient Descent (SGD) optimizer were used.
- Baseline HTK system: A similar GMM-HMM model with 8 GMMs assigned to each state. The Baum-Welch reestimation algorithm is used for training.

Performance Evaluation:

- Epoch scoring with an epoch duration of 1 sec is used.
- The LDA-MLLT system performs the best among all Kaldi systems.
- Kaldi performance is lacking compared to a GMM-HMM model of HTK system.
- SPSW event is consistently harder for all recognition systems to detect.

Ref/Hyp	PLED	GPED	SPSW	ARTF	EYEM	BCKG
PLED	46.68	18.19	22.22	1.58	7.39	3.94
GPED	41.00	20.36	8.94	6.21	12.35	11.14
SPSW	6.46	34.26	11.41	2.95	16.29	28.62
ARTF	2.59	1.41	0.76	13.76	40.19	41.30
EYEM	2.80	0.26	0.33	5.62	48.90	42.09
BCKG	5.38	5.26	8.97	1.93	10.78	67.68

Kaldi's GMM-HMM Monophone System

Ref/Hyp	PLED	GPED	SPSW	ARTF	EYEM	BCKG
PLED	43.08	26.87	11.11	5.31	11.38	2.26
GPED	29.80	31.01	5.92	10.58	3.98	18.71
SPSW	1.34	37.12	3.33	21.69	18.64	17.88
ARTF	0.63	2.52	0.72	30.67	38.43	27.03
EYEM	0.22	1.92	0.00	8.60	54.36	34.89
BCKG	2.29	1.52	10.21	15.21	8.69	62.08

Kaldi's LDA-MLLT Triphone System

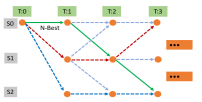
Ref/Hyp	PLED	GPED	SPSW	ARTF	EYEM	BCKG
PLED	46.83	28.38	9.58	3.96	3.52	7.70
GPED	21.20	20.87	2.93	43.78	0.39	10.80
SPSW	0.04	36.55	1.80	9.81	9.7	42.04
ARTF	0.39	2.58	6.15	24.35	27.87	38.65
EYEM	0.42	0.35	3.63	1.70	25.75	68.11
BCKG	0.35	2.18	7.63	13.31	2.71	73.79

Kaldi's DNN-HMM MLP System

Ref/Hyp	PLED	GPED	SPSW	ARTF	EYEM	BCKG
PLED	54.80	17.62	20.32	0.70	4.70	1.85
GPED	20.42	53.32	11.55	7.46	2.39	4.85
SPSW	4.59	33.33	40.21	4.41	9.17	8.29
ARTF	2.77	3.81	4.81	41.24	2.18	45.19
EYEM	8.51	2.13	4.26	2.13	82.37	0.61
BCKG	2.28	7.37	8.81	7.02	2.59	71.93

HTK's Baseline GMM-HMM System

Decoding



- Lattices are used during decoding to track temporal information.
- Each node of a lattice includes acoustic as well as language model scores.
- The Viterbi algorithm is used during decoding with the lattice beam value within the range (0.7 – 1.0).
- Decoding was performed using 1-best, pushed lattices and event-level posteriors. Event-level posteriors gives the most balanced results.
- Event-level posteriors are collected for each frame and mapped back to its time-stamp/phone boundary in the output hypothesis files.

Parameter Count and Complexity

- Kaldi's baseline GMM-HMM monophone model uses 8,438 parameters. A similar model based on HTK uses 12,494 parameters.
- Kaldi's Viterbi training requires ~40 minutes to train the model across 8 CPU cores. HTK required same amount of time for Baum-Welch training on only 1 CPU core.

Summary

- EEGs show similar behavior to speech waveforms but certain morphologies such as spike / sharp and wave discharges are difficult to detect.
- Kaldi's multi-pass HMMs developed for the six-way classification do not show any improvement over our HTK baseline system.
- LDA-MLLT system's perform better than Kaldi's other variants. However, systems which use LDA features perform extremely poorly on SPSW events.

Future Work

- Focus on the development of Kaldi's complex DNN variants on this data. i.e. T-DNN and Kaldi's nnet2, nnet3 recipes.
- Investigate Kaldi's approaches to speech activity detection and integrate more temporal information about the waveshapes.
- Implement neural network postprocessors to learn the spatial context of the signals along with its temporal properties.

Acknowledgements

- Research reported in this poster was supported by National Human Genome Research Institute of the National Institutes of Health under award number 3U01HG008468-02S1. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
- The authors would like to thank Dan Povey, Alan McCree, John Steinberg and many other members of the Kaldi and HLTCoE teams for their guidance in adapting Kaldi to this task.